

VALIDAÇÃO DE ALGORITMOS DE IA EM TRIAGEM POPULACIONAL OFTALMOLÓGICA

VALIDATION OF AI ALGORITHMS IN OPHTHALMOLOGICAL POPULATION SCREENING

GABRIEL RIBEIRO DE OLIVEIRA SILVA¹, ÍTALO BEZERRA ALVES², ISABELA LUIZA PACHECO RODRIGUES²,
DOUGLAS WALLACY DA SILVA AMÂNCIO², FABIANO INÁCIO DE SOUZA³

1. Graduando em Medicina na Universidade Federal de Jataí, Jataí/GO, Brasil.
2. Graduando em Medicina na Universidade Federal de Goiás, Goiânia/GO, Brasil.
3. Médico Ortopedista. Doutor em Ortopedia, Traumatologia e Reabilitação, Goiânia/GO, Brasil.

RESUMO

A aplicação da inteligência artificial (IA) na oftalmologia tem despontado como uma estratégia promissora para o diagnóstico precoce e a triagem populacional de doenças visuais, mitigando a escassez de especialistas e otimizando a gestão de recursos em saúde. O presente estudo tem como objetivo analisar a eficácia clínica, a robustez e a capacidade de generalização de algoritmos de IA aplicados ao rastreamento e apoio diagnóstico de patologias oftalmológicas. Para isso, conduziu-se uma revisão integrativa da literatura através de buscas nas bases de dados Google Acadêmico, PubMed, ResearchGate e Portal de Periódicos CAPES, culminando na seleção e análise de dez trabalhos publicados entre 2018 e 2025. A síntese das evidências demonstrou que os modelos baseados em aprendizagem profunda (deep learning), voltados para a análise de imagens do fundo do olho, apresentam altos índices de sensibilidade e especificidade, sobretudo na detecção de retinopatia diabética, retinopatia da prematuridade, glaucoma e degeneração macular relacionada com a idade. Constatou-se, ainda, que a integração destas tecnologias amplia significativamente a capacidade de rastreamento clínico. Contudo, o desempenho destes algoritmos pode sofrer variações significativas perante populações distintas ou diferentes equipamentos de captura, um desafio técnico denominado domain shift. Infere-se, portanto, que a implementação clínica segura da IA requer rigorosas validações externas multicêntricas e a elaboração de diretrizes ético-regulatórias claras, atuando como um complemento ao julgamento médico e democratizando o acesso aos cuidados visuais.

Palavras-chave: Inteligência artificial, Oftalmologia, Triagem, Algoritmos, Validação externa.

ABSTRACT

The application of artificial intelligence (AI) in ophthalmology has emerged as a promising strategy for the early diagnosis and population screening of visual diseases, mitigating the shortage of specialists and optimizing the management of health resources. The present study aims to analyze the clinical efficacy, robustness, and generalization capacity of AI algorithms applied to the screening and diagnostic support of ophthalmological

pathologies. For this purpose, an integrative literature review was conducted through searches in the Google Scholar, PubMed, ResearchGate, and CAPES Periodicals Portal databases, culminating in the selection and analysis of ten studies published between 2018 and 2025. The synthesis of evidence demonstrated that deep learning models focused on the analysis of fundus images present high rates of sensitivity and specificity, especially in the detection of diabetic retinopathy, retinopathy of prematurity, glaucoma, and age-related macular degeneration. Furthermore, it was found that the integration of these technologies significantly expands clinical screening capacity. However, the performance of these algorithms can undergo significant variations when faced with distinct populations or different capture equipment, a technical challenge known as domain shift. Therefore, it is inferred that the safe clinical implementation of AI requires rigorous multicenter external validations and the development of clear ethical-regulatory guidelines, acting as a complement to medical judgment and democratizing access to visual care.

Keywords: Artificial intelligence, Ophthalmology, Screening, Algorithms, External validation.

INTRODUÇÃO

As doenças oftalmológicas representam um desafio substancial para a saúde pública global, sendo responsáveis por altas taxas de morbidade e perda visual evitável. Em cenários marcados pela escassez de médicos especialistas e por barreiras geográficas ou socioeconômicas, o acesso oportuno ao diagnóstico e ao tratamento precoce torna-se limitado, sobrecarregando os sistemas de saúde e agravando o prognóstico dos pacientes.¹

Neste contexto, a integração de novas tecnologias tem impulsionado a criação de estratégias inovadoras para a triagem populacional. A inteligência artificial (IA), especialmente através de algoritmos de aprendizagem profunda (deep learning), tem emergido como uma força transformadora na oftalmologia contemporânea.² A utilização destas ferramentas associada à teleoftalmologia tem demonstrado grande capacidade para analisar fotografias do fundo do olho e estratificar riscos com alta precisão, otimizando o fluxo assistencial em patologias altamente prevalentes, como a retinopatia diabética, o glaucoma e a degeneração macular relacionada com a idade.^{3,4}

Apesar dos resultados promissores alcançados durante a fase de desenvolvimento em ambientes controlados, a literatura científica levanta questões sobre a segurança e a eficácia destas ferramentas em cenários reais. O desempenho dos algoritmos sofre, muitas vezes, um declínio quando exposto a novas populações, diferentes protocolos de aquisição de imagens ou variações de equipamentos, constituindo um desafio técnico reconhecido como domain shift.⁵

Portanto, a compreensão das limitações e da acuidade destas tecnologias assume particular relevância. Avaliar a robustez dos modelos de IA por via de processos de validação externa torna-se essencial para garantir que a sua incorporação em programas de rastreio em larga escala ocorra de forma segura, fiável e equitativa, sem comprometer a qualidade dos cuidados prestados aos doentes.^{6,7}

OBJETIVOS

Objetivo geral:

- Analisar a robustez, o desempenho diagnóstico e a capacidade de generalização de algoritmos de inteligência artificial direcionados para a triagem populacional de patologias oftalmológicas, a partir das evidências científicas compiladas na literatura.

Objetivos específicos:

- Classificar os modelos de aprendizagem profunda (deep learning) e as respectivas patologias oculares

mapeadas nos estudos integrados, traçando o panorama atual das ferramentas tecnológicas aplicadas à saúde visual.

- Distinguir a variabilidade dos indicadores de acuidade diagnóstica (sensibilidade, especificidade e área sob a curva ROC) das ferramentas de IA em função de diferentes patologias, como a retinopatia diabética, o glaucoma, a retinopatia da prematuridade e a degeneração macular relacionada com a idade.
- Enumerar as principais limitações metodológicas e os desafios ético-técnicos reportados, com especial enfoque no fenômeno de desvio de domínio (domain shift) e na dependência técnica da qualidade das imagens captadas.
- Selecionar as perspectivas de aplicabilidade clínica futura e o impacto em saúde pública decorrentes da integração destes algoritmos com a atenção primária e com os sistemas de teleoftalmologia.

MÉTODOS

O presente artigo caracteriza-se como uma revisão integrativa da literatura, de natureza qualitativa e exploratória, focada na validação de algoritmos de Inteligência Artificial (IA) aplicados à triagem populacional de patologias oftalmológicas. A coleta dos dados foi realizada em maio de 2026, por meio de consulta sistemática às bases Google Acadêmico, PubMed, ResearchGate e Portal de Periódicos CAPES, selecionadas pela abrangência de periódicos de alto impacto e pelo acesso a produções científicas recentes sobre diagnóstico em saúde visual.

A estratégia de busca empregou descritores controlados e termos técnicos em português e inglês, combinados com operadores booleanos básicos ("Inteligência Artificial" AND "Oftalmologia", "AI algorithms" AND "Screening" AND "Retina"), de modo a garantir a intersecção entre tecnologia e prática clínica, bem como o agrupamento de variações terminológicas como redes neurais profundas e aprendizado de máquina.

Os critérios de inclusão compreenderam artigos originais e revisões publicados entre 2018 e 2025, que abordassem especificamente a acurácia, sensibilidade e especificidade de ferramentas de IA em contextos de triagem. Após a coleta inicial, títulos e resumos foram analisados para descartar estudos duplicados ou sem dados validados de desempenho algorítmico. Os textos selecionados foram lidos na íntegra, permitindo a síntese dos principais modelos de validação externa e das evidências sobre a viabilidade dessas tecnologias em sistemas públicos de saúde e programas de rastreamento em massa.

RESULTADOS E DISCUSSÕES

• Caracterização dos estudos incluídos

Foram incluídos 10 estudos publicados entre 2018 e 2025, abrangendo estudos originais de validação diagnóstica, estudos observacionais e revisões de literatura. As principais características dos trabalhos incluídos encontram-se na Tabela 1. Observou-se predomínio de pesquisas voltadas à aplicação e validação de algoritmos de inteligência artificial (IA) em doenças da retina, com menor frequência de outras condições oftalmológicas.

Tabela 1: Caracterização dos estudos incluídos.

Autor/ano	Tipo de estudo	Validou o algoritmo?	País/cenário
Pionório et al., 2024	Revisão integrativa	N/A	N/A
Checon et al., 2025	Pesquisa qualitativa/exploratória	N/A	N/A
Mamprin et al., 2025	Revisão narrativa	N/A	Atenção primária à saúde
Arenas-Cavalli et al., 2021	Estudo observacional de validação clínica	Sim	Chile; sistema nacional de saúde
Han et al., 2022	Estudo prospectivo de validação	Sim	China; hospital e comunidade
Coyner et al., 2022	Estudo diagnóstico com validação externa multicêntrica	Sim	Índia, Nepal e Mongólia; programa de telemedicina
Zapata et al., 2020	Estudo retrospectivo	Sim	Espanha; teleoftalmologia
Kuiava et al., 2021	Estudo de desenvolvimento/validação	Sim	Brasil; teleoftalmologia
Chen et al., 2025	Estudo diagnóstico com workflow	Sim	EUA e Singapura
Chen et al., 2021	Estudo observacional de desenvolvimento/validação	Sim	Taiwan

Fonte: Pionório et al., 2024; Checon et al., 2025; Mamprin et al., 2025; Arenas-Cavalli et al., 2021; Han et al., 2022; Coyner et al., 2022; Zapata et al., 2020; Kuiava et al., 2021; Chen et al., 2025; Chen et al., 2021.^{3,4,8-14}

Tabela 2: Descrição e resultados dos estudos

Autor/ano	Doença/aplicação	Modelo de IA	Principal contribuição
Pionório et al., 2024	Foco em retinopatia diabética e glaucoma	IDx-DR, Eyer e outros sistemas	Reforça o papel na democratização do diagnóstico e no apoio clínico
Checon et al., 2025	Foco em retinopatia diabética, glaucoma e degeneração macular relacionada à idade	Síntese de estudos sobre deep learning	Destaca baixa diversidade dos bancos de dados e necessidade de validação local
Mamprin et al., 2025	Triagem de retinopatia diabética	Deep learning	Alta performance diagnóstica e potencial de reduzir custos
Arenas-Cavalli et al., 2021	Triagem de retinopatia diabética	Sistema DART com redes convolucionais	Sensibilidade 94,6%, especificidade 74,3%; viabilidade de aplicação a rede pública heterogênea
Han et al., 2022	Retinopatia diabética referenciável, doença macular reticular e glaucoma suspeito	AI-100 (deep learning)	Robustez em cenários reais, porém com desvio de domínio entre câmeras e populações
Coyner et al., 2022	Triagem de retinopatia da prematuridade	Baseado em idade gestacional e score de gravidade vascular (VSS)	Sensibilidade de 100% nos três cenários e redução estimada de 38–51% dos exames
Zapata et al., 2020	Identificação de qualidade, lateralidade, degeneração macular relacionada à idade e glaucoma suspeito	5 algoritmos de deep learning / rede neural convolucional	Alta acurácia para múltiplas tarefas; útil para o fluxo de triagem
Kuiava et al., 2021	Glaucoma, retinopatia diabética, toxoplasmose ocular, papiledema, descolamento de retina, retina normal	MobileNet e Inception	Sensibilidade ~91% e especificidade ~98%; útil como ferramenta de apoio e triagem
Chen et al., 2025	Diagnóstico/classificação de degeneração macular relacionada à idade	DeepSeeNet e DeepSeeNet+	Aumentou a acurácia de 23/24 clínicos; Reduziu tempo diagnóstico; reforça necessidade de accountability
Chen et al., 2021	Medidas palpebrais para ptose: MRD1, MRD2 e LF	MAIA software e redes convolucionais	Mostra aplicação de IA via smartphone para mensuração objetiva e rápida

Fonte: Pionório et al., 2024; Checon et al., 2025; Mamprin et al., 2025; Arenas-Cavalli et al., 2021; Han et al., 2022; Coyner et al., 2022; Zapata et al., 2020; Kuiava et al., 2021; Chen et al., 2025; Chen et al., 2021.^{3,4,8-14}

De modo geral, houve predominância de algoritmos baseados em deep learning, aplicados à análise de fotografias de fundo de olho, com foco principal em retinopatia diabética, retinopatia da prematuridade, degeneração macular relacionada à idade (AMD) e glaucoma suspeito.⁸⁻¹⁴ Apesar das diferenças metodológicas observadas, os trabalhos convergiram na avaliação do potencial da IA para apoiar processos de triagem e ampliar o acesso ao cuidado oftalmológico, frequentemente em cenários de teleoftalmologia e de rastreamento em larga escala.^{4,8-11} Paralelamente, as revisões incluídas reforçaram essas tendências gerais, principalmente em relação às perspectivas de incorporação dessas ferramentas em programas de saúde visual.^{3,4}

• Desempenho dos algoritmos de IA

Os estudos analisados demonstraram desempenho global elevado para os algoritmos de IA, comportamento evidenciado pelos altos valores de sensibilidade, especificidade e área sob a curva ROC, reforçando o potencial dessas ferramentas para uso em programas de rastreamento e de apoio à decisão clínica.⁹⁻¹⁴

A retinopatia diabética (RD) foi a condição em que esse desempenho se mostrou mais estável. Em um contexto de atenção primária e triagem populacional de RD⁸, sistemas baseados em aprendizado profundo foram descritos com acurácia superior a 90%, acompanhada de sensibilidade e especificidade em torno de 95%, além de potencial redução dos custos do rastreio feito manualmente.⁸ Já na validação clínica conduzida no Chile⁹, o sistema DART apresentou sensibilidade de 94,6%, especificidade de 74,3% e AUC de 0,915 em 1123 exames de olhos diabéticos, sugerindo aplicabilidade em uma rede nacional heterogênea⁹. Enquanto isso, em um cenário real de comunidade e hospital na China¹⁰, o AI-100 também demonstrou desempenho robusto para retinopatia diabética referenciável, com sensibilidade de 95%, especificidade de 96,3% e AUC de 0,954.¹⁰ Desse modo, os achados evidenciam que a IA apresenta elevado potencial para triagem de RD, principalmente em programas populacionais e na teleoftalmologia.

Na retinopatia da prematuridade, o destaque foi a capacidade de antecipação do risco. O modelo baseado em idade gestacional e vascular severity score¹¹ manteve desempenho consistente nas populações avaliadas, com sensibilidade de 100% nos três coortes, especificidade de 63,3% na Índia, 77,8% no Nepal e 45,8% na Mongólia, além de AUC de 0,919 na validação cruzada de 5 partes e 0,944 no conjunto de validação. Esse desempenho permitiu identificar com antecedência os casos que evoluíram para necessidade de tratamento, reforçando o papel da IA não apenas como ferramenta diagnóstica, mas também como recurso preditivo em um cenário no qual a detecção precoce tem impacto direto no prognóstico visual.¹¹

Os estudos que tiveram avaliação da IA voltada ao glaucoma ou ao glaucoma suspeito mostraram desempenho favorável, embora mais heterogêneo do que o observado nas doenças retinianas.^{10,12,13} Em uma base espanhola com mais de 300 mil imagens¹², os modelos de deep learning alcançaram AUC de 0,863, acurácia de 80,2%, sensibilidade de 76,8% e especificidade de 83,8% para neuropatia glaucomatosa suspeita, mantendo aplicabilidade para rastreamento, ainda que com rendimento inferior ao de outras doenças analisadas no mesmo trabalho. Em contrapartida, o algoritmo AI-10010 apresentou desempenho superior em um contexto clínico real, com sensibilidade de 95%, especificidade de 94,6% e AUC de 0,976 para glaucoma suspeito, em um ambiente de validação com diferentes câmeras e cenários assistenciais. Resultados igualmente elevados foram observados no software brasileiro de triagem¹³, cujos modelos MobileNet e Inception atingiram sensibilidades entre 93,33% e 96,55% e especificidades acima de 99% para

glaucoma. A avaliação conjunta dessas pesquisas sugere que a IA possui potencial significativo para triagem de glaucoma, embora a performance varie conforme a qualidade das imagens, as características populacionais e a composição dos bancos de dados.^{10,12,13}

Na AMD, os resultados também foram consistentes, tanto para detecção da doença quanto para apoio ao especialista.^{12,14} No estudo realizado na Espanha¹², modelos de deep learning alcançaram AUC de 0,936, acurácia de 86,3%, sensibilidade de 90,2% e especificidade de 82,5%, apontando para uma distinção eficaz entre indivíduos acometidos e não acometidos. Em um trabalho multicêntrico, foi avaliado um fluxo assistido por IA para diagnóstico e classificação da gravidade da AMD¹⁴. Nele, a assistência da IA elevou o F1-score médio dos clínicos de 37,71 para 45,52 e reduziu o tempo estimado de diagnóstico, embora o desempenho inferior observado na coorte externa de Singapura sugira possíveis limitações de generalização mesmo em modelos promissores. Assim, a AMD aparece como uma condição em que a IA pode apoiar tanto a detecção quanto a estratificação de gravidade, mas ainda depende de refinamento para uso ampliado em populações mais diversas.^{12,14}

Além das aplicações voltadas a doenças específicas, alguns estudos evidenciaram o potencial da IA para atuação em cenários oftalmológicos variados.^{13,15} O software brasileiro desenvolvido para classificação automatizada de doenças do fundo de olho¹³ demonstrou capacidade de identificação de outros cenários oftalmológicos para além do glaucoma e da retinopatia diabética, incluindo toxoplasmose ocular, papiledema, descolamento de retina e imagens normais, alcançando sensibilidade global superior a 91% e especificidade próxima de 99% para as arquiteturas MobileNet e Inception. De forma semelhante, um sistema baseado em fotografias obtidas por smartphone para mensuração automatizada de parâmetros palpebrais¹⁵ apresentou elevada concordância com as avaliações realizadas por especialistas, com correlações elevadas para MRD1, MRD2 e função do músculo levantador. Somados, esses resultados sugerem que a IA pode atuar não apenas na triagem de doenças retinianas, mas também como ferramenta multimodal de apoio diagnóstico, telemedicina e padronização objetiva de medidas clínicas em oftalmologia.^{13,15}

• Validação externa e generalização dos modelos

A validação externa emergiu como um dos principais critérios para avaliar a aplicabilidade clínica dos algoritmos de IA.^{4,14} Os estudos analisados mostraram que indicadores elevados de desempenho obtidos durante o desenvolvimento não garantem, por si só, resultados equivalentes quando os modelos são aplicados a populações distintas ou expostos a fatores como diferenças epidemiológicas, qualidade das imagens, protocolos de aquisição e variações entre equipamentos.^{9,10} Nesse contexto, mesmo algoritmos com desempenho elevado em suas bases de treinamento podem apresentar redução de performance em cenários externos, fenômeno conhecido como domain shift.⁵ Assim, a robustez, a confiabilidade e a capacidade de adaptação mostram-se tão importantes quanto o desempenho diagnóstico em si^{4,14}, reforçando a necessidade de validações independentes e multicêntricas antes da adoção em larga escala, especialmente em programas de rastreamento populacional e sistemas públicos de saúde.⁷

Os trabalhos incluídos empregaram diferentes estratégias de validação externa, com destaque para validações multicêntricas em populações independentes^{11,14}, análises em distintos níveis de atenção à saúde¹⁰ e testes realizados com diferentes equipamentos de captura de imagem.^{9,10} Em menor frequência, houve também avaliações voltadas à integração dos modelos em fluxos clínicos reais, não apenas para verificar a performance diagnóstica, mas também sua utilidade

prática.¹⁴ Somadas, essas abordagens permitiram avaliar a capacidade dos algoritmos de manter desempenho e aplicabilidade fora do ambiente em que foram originalmente desenvolvidos.^{10,11,14}

Entre as estratégias empregadas, as validações multicêntricas forneceram as evidências mais robustas de generalização, demonstrando manutenção do desempenho mesmo diante de diferenças epidemiológicas, operacionais e tecnológicas.^{10,11} Entretanto, a redução da performance observada em algumas coortes externas indica que a generalização ainda representa um desafio relevante para parte dos modelos avaliados.¹⁴ Esses achados reforçam que a validação externa deve ser compreendida como um requisito fundamental para a incorporação segura da IA em programas de triagem oftalmológica.^{4,14,16}

• Aplicabilidade clínica e impacto em saúde pública

Os estudos incluídos demonstraram que a principal contribuição clínica da IA para a oftalmologia está relacionada à ampliação da capacidade de triagem e à qualificação dos fluxos assistenciais.^{3,4,8} Em termos práticos, os algoritmos mostraram potencial para identificar precocemente indivíduos com maior risco de doença ocular, o que acaba por favorecer o encaminhamento oportuno e reduzir a sobrecarga dos serviços de referência.⁹⁻¹¹ Esse potencial foi observado sobretudo em programas de rastreamento de doenças de alta prevalência, nos quais a IA foi frequentemente proposta como ferramenta complementar à atenção primária e à teleoftalmologia.^{4,8,9,17}

Outro aspecto recorrente foi a racionalização dos recursos em saúde.^{4,8} As pesquisas sugerem que a incorporação de sistemas de IA pode contribuir para reduzir exames desnecessários e melhorar a alocação da força de trabalho especializada.^{8,11} Em alguns cenários, observou-se potencial de diminuição da carga assistencial, porém sem comprometer a sensibilidade diagnóstica, o que posiciona essas tecnologias como instrumentos de apoio à gestão dos programas de rastreamento.^{10,11,14,18,19}

A relevância em saúde pública foi particularmente evidente em contextos marcados pela escassez de oftalmologistas e por dificuldades de acesso aos serviços especializados.^{3,4,11} Nesse sentido, a associação entre IA, teleoftalmologia e plataformas digitais foi apontada como uma estratégia capaz de ampliar a cobertura diagnóstica e fortalecer programas de prevenção, especialmente em regiões com barreiras geográficas de acesso.^{3,4,9} Em populações vulneráveis e em sistemas de saúde com recursos limitados, esses algoritmos surgem como uma via concreta para expandir o cuidado oftalmológico.^{1,4,8,11,17}

• Limitações metodológicas e desafios ético-técnicos

Apesar dos resultados promissores dos estudos incluídos, diversas limitações metodológicas foram identificadas e devem ser consideradas na interpretação dos achados. A principal delas foi a heterogeneidade entre os trabalhos, os quais diferiram quanto às doenças avaliadas, características das populações, estruturas metodológicas, equipamentos utilizados e métricas de desempenho reportadas.^{3,4} Além disso, os algoritmos foram desenvolvidos para finalidades distintas, como triagem, classificação de gravidade, predição de risco e apoio à decisão clínica, o que dificulta comparações diretas e limita a padronização das evidências disponíveis.^{3,4,20}

Outro ponto crítico foi a dependência dos algoritmos em relação à qualidade dos dados utilizados.^{9,10,11} Os estudos destacaram que fatores como qualidade das imagens, diferenças entre equipamentos, protocolos de aquisição e características específicas das populações avaliadas podem influenciar significativamente o desempenho dos modelos.^{9-11,14} Essa variabilidade favorece a ocorrência do domain shift, evidenciando que a confiabilidade dos algoritmos permanece diretamente relacionada à

qualidade e à representatividade dos dados empregados em seu treinamento e aplicação.^{10,14,21}

Na dimensão ética e regulatória, os trabalhos apontaram preocupações relacionadas à interpretabilidade dos modelos, à diversidade limitada de determinadas bases de treinamento, ao risco de vieses algorítmicos e à definição de responsabilidades clínicas em situações de erro de diagnóstico.^{4,14} Também foram ressaltadas questões ligadas à governança dos dados, à transparência dos sistemas e à necessidade de supervisão profissional contínua.^{4,14,22,23} Dessa forma, a implementação segura dessas tecnologias depende não apenas do desempenho técnico dos algoritmos, mas também de mecanismos de regulação e monitoramento que assegurem sua segurança, confiabilidade e uso equitativo em diferentes contextos de aplicação clínica.^{4,14,24}

• Perspectivas futuras

As perspectivas futuras apontadas pelos trabalhos convergem para a consolidação da IA como ferramenta integrada aos fluxos assistenciais oftalmológicos^{2-4,14}, o que exige avanços relacionados à robustez, generalização e validação dos algoritmos.^{4,14} Nesse sentido, destacam-se como prioridades a realização de estudos prospectivos, multicêntricos e conduzidos em populações mais diversas, além da ampliação da representatividade dos bancos de dados utilizados para treinamento e validação.^{4,11,14} Essas medidas são essenciais para aumentar a confiabilidade dos sistemas e reduzir limitações relacionadas ao domain shift e aos vieses de seleção, especialmente quando se pretende incorporar esses modelos a diferentes contextos de cuidado em saúde.^{4,6,10,14}

Outra tendência recorrente é a integração entre inteligência artificial, teleoftalmologia e fluxos clínicos já estabelecidos.^{3,4,14} Os estudos sugerem que o futuro dessas tecnologias está menos relacionado à substituição do especialista e mais ao desenvolvimento de ferramentas capazes de apoiar a tomada de decisão, otimizar programas de rastreamento e ampliar o acesso ao cuidado oftalmológico.^{4,8,14,25} Assim, aspectos como interpretabilidade dos algoritmos, monitoramento contínuo do desempenho e integração segura aos sistemas de saúde tendem a ocupar papel central na consolidação da IA como estratégia complementar na prevenção, detecção e manejo das condições oftalmológicas.^{4,14,26,27}

CONCLUSÃO

A revisão evidenciou que os algoritmos de inteligência artificial, especialmente aqueles baseados em deep learning aplicados à análise de imagens oftalmológicas, apresentam desempenho elevado para triagem e apoio diagnóstico de diversas doenças oculares. Os estudos conduzidos sobre retinopatia diabética, retinopatia da prematuridade, degeneração macular relacionada à idade e glaucoma demonstraram, de modo geral, altos índices de sensibilidade, especificidade e área sob a curva ROC. Dentre essas doenças, o modelo de triagem populacional e aplicação na atenção primária na detecção de retinopatia diabética obteve os melhores resultados dentre as doenças analisadas, com altas taxas de acurácia, especificidade e sensibilidade. Tal análise reforça o potencial dessas ferramentas para ampliar a capacidade de rastreamento, otimizar fluxos assistenciais e favorecer a detecção precoce de condições associadas à perda visual evitável.

Entretanto, os achados também evidenciam que o desempenho obtido exclusivamente em bases de desenvolvimento não garante a aplicabilidade clínica dos modelos, visto que fatores limitantes como qualidade das imagens, diversidade, tamanho dos bancos de dados e diferenças entre equipamentos tecnológicos utilizados são desafios relevantes para a implementação segura dessas tecnologias. Nesse sentido, a validação externa mostrou-se como uma etapa fundamental para avaliar a robustez,

a confiabilidade e a capacidade de generalização dos algoritmos em diferentes populações, contextos epidemiológicos e condições operacionais.

Além do elevado potencial diagnóstico, a IA se mostrou uma ferramenta com grande capacidade de contribuir para a ampliação do acesso ao cuidado oftalmológico, especialmente quando integrada à teleconsulta oftalmológica e a programas de rastreamento em larga escala, podendo beneficiar populações com acesso limitado a especialistas. Contudo, sua incorporação efetiva aos sistemas de saúde exige o enfrentamento de questões metodológicas, éticas e regulatórias relacionadas à transparência dos modelos, mitigação de vieses, supervisão clínica, regulamentação e segurança dos dados.

Portanto, a inteligência artificial representa uma ferramenta promissora e potencialmente transformadora para a oftalmologia contemporânea. Todavia, o futuro da IA na especialidade não aponta para a substituição do julgamento clínico, mas sim como um instrumento de auxílio para a melhor eficiência diagnóstica.

REFERÊNCIAS

1. Kawaguchi A, Sharafeldin N, Sundaram A, Campbell S, Tennant M, Rudnisky C, Weis E, Damji KF. Tele-Ophthalmology for Age-Related Macular Degeneration and Diabetic Retinopathy Screening: A Systematic Review and Meta-Analysis. *Telemed J E Health*. 2018; 24(4): 301-308. doi: 10.1089/tmj.2017.0100.
2. Jin K, Ye J. Artificial intelligence and deep learning in ophthalmology: Current status and future perspectives. *Adv Ophthalmol Pract Res*. 2022; 2(3): 100078. doi: 10.1016/j.aopr.2022.100078.
3. Pionório VAL, Silva DF da, Pionório JL. Avanços recentes da IA na oftalmologia - revisão de literatura. *Braz J Health Rev [Internet]*. 2024 May 15 [Citado em 07 Jun 2026]; 7(3): 1-19. Disponível em: <https://ojs.brazilianjournals.com.br/ojs/index.php/BJHR/article/view/69712>.
4. Checon MES, Grill HM, Wendt MG, Nunes GT, Ávila GC de, Ferreira NM, Carvalho GSC, Andrezza IC, Rocha FLA da. Inteligência artificial na oftalmologia: Diagnóstico preciso, acesso ampliado e novos desafios. *JMBR [Internet]*. 2025 Aug 9 [Citado em 07 Jun 2026]; 2(4): 771-779. Disponível em: <https://journalmbr.com.br/index.php/jmbr/article/view/831>.
5. Guan H, Liu M. Domain Adaptation for Medical Image Analysis: A Survey. *IEEE Trans Biomed Eng*. 2022; 69(3): 1173-1185. doi: 10.1109/TBME.2021.3117407.
6. Rivera SC, Liu X, Chan AW, Denniston AK, Calvert MJ. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI Extension. *BMJ*. 2020; 370: m3210. doi: 10.1136/bmj.m3210.
7. Sounderajah V, Guni A, Liu X, Collins GS, Karthikesalingam A, Markar SR, Golub RM, Denniston AK, Shetty S, Moher D, Bossuyt PM, Darzi A, Ashrafian H. The STARD-AI reporting guideline for diagnostic accuracy studies using artificial intelligence. *Nat Med*. 2025; 31(10): 3283-3289. doi: 10.1038/s41591-025-03953-8.
8. Mamprin J, Prates NA, Mascarenhas Júnior OMR, Rossoni LT, Massote G de B. Inteligência artificial na triagem da retinopatia diabética na atenção primária: acurácia diagnóstica e custo-efetividade. *REASE [Internet]*. 2025 Nov 10 [Citado em 07 Jun 2026]; 11(11): 2177-2187. Disponível em: <https://periodicorease.pro.br/rease/article/view/22147>. doi: 10.51891/rease.v11i11.22147.
9. Arenas-Cavalli JT, Abarca I, Rojas-Contreras M, Bernuy F, Donoso R. Clinical validation of an artificial intelligence-based diabetic retinopathy screening tool for a national health system. *Eye (Lond)*. 2022; 36(1): 78-85. doi: 10.1038/s41433-020-01366-0.
10. Han R, Cheng G, Zhang B, Yang J, Yuan M, Yang D, Wu J, Liu J, Zhao C, Chen Y, Xu Y. Validating automated eye disease screening AI algorithm in community and in-hospital scenarios. *Front Public Health*. 2022; 10: 944967. doi: 10.3389/fpubh.2022.944967.
11. Coyner AS, Oh MA, Shah PK, Singh P, Ostmo S, Valikodath NG, Cole E, Al-Khaled T, Bajimaya S, K C S, Chuluunbat T, Munkhuu B, Subramanian P, Venkatapathy N, Jonas KE, Hallak JA, Chan RVP, Chiang MF, Kalpathy-Cramer J, Campbell JP. External Validation of a Retinopathy of Prematurity Screening Model Using Artificial Intelligence in 3 Low- and Middle-Income Populations. *JAMA Ophthalmol*. 2022; 140(8): 791-798. doi: 10.1001/jamaophthalmol.2022.2135.
12. Zapata MA, Royo-Fibla D, Font O, Vela JJ, Marcantonio I, Moya-Sánchez EU, Sánchez-Pérez A, García-Gasulla D, Cortés U, Ayguadé E, Labarta J. Artificial Intelligence to Identify Retinal Fundus Images, Quality Validation, Laterality Evaluation, Macular Degeneration, and Suspected Glaucoma. *Clin Ophthalmol*. 2020; 14: 419-429. doi: 10.2147/OPTH.S235751.
13. Kuiava VA, Kuiava EL, Chielle EO, Syllós R. Desenvolvimento de sistema estruturado com inteligência artificial para apoio no diagnóstico de patologias oftalmológicas mais relevantes. *Clin Biomed Res [Internet]*. 2021 Jun 28 [Citado em 07 Jun 2026]; 41(1): 27-32. Disponível em: <https://seer.ufrgs.br/index.php/hcpa/article/view/109565>.
14. Chen Q, Keenan TDL, Agron E, Allot A, Guan E, Duong B, Elsayy A, Hou B, Xue C, Bhandari S, Broadhead G, Cousineau-Krieger C, Davis E, Gensheimer WG, Golshani CA, Grasic D, Gupta S, Haddock L, Konstantinou E, Lamba T, Maiberger M, Mantopoulos D, Mehta MC, Elnahry AG, Al-Nawafli M, Oshinsky A, Powell BE, Purt B, Shin S, Stiefel H, Thavikulwat AT, Wroblewski KJ, Tham YC, Cheung CMG, Cheng CY, Chew EY, Hribar MR, Chiang MF, Lu Z. AI Workflow, External Validation, and Development in Eye Disease Diagnosis. *JAMA Netw Open*. 2025; 8(7): e2517204. doi: 10.1001/jamanetworkopen.2025.17204.

15. Chen HC, Tzeng SS, Hsiao YC, Chen RF, Hung EC, Lee OK. Smartphone-Based Artificial Intelligence-Assisted Prediction for Eyelid Measurements: Algorithm Development and Observational Validation Study. *JMIR Mhealth Uhealth*. 2021; 9(10): e32444. doi: 10.2196/32444.
16. Alqahtani AS, Alshareef WM, Aljadani HT, Hawsawi WO, Shaheen MH. The efficacy of artificial intelligence in diabetic retinopathy screening: a systematic review and meta-analysis. *Int J Retina Vitreous*. 2025; 11(1): 48. doi: 10.1186/s40942-025-00670-9.
17. Kalogeropoulos D, Kalogeropoulos C, Stefanidou M, Neofytou M. The role of tele-ophthalmology in diabetic retinopathy screening. *J Optom*. 2020 Oct-Dec;13(4):262-268. doi: 10.1016/j.optom.2019.12.004. Epub 2020 Jan 14. PMID: 31948924; PMCID: PMC7520530.
18. El Arab RA, Al Moosa OA. Systematic review of cost effectiveness and budget impact of artificial intelligence in healthcare. *NPJ Digit Med*. 2025; 8(1): 548. doi: 10.1038/s41746-025-01722-y.
19. Abramoff MD, Whitestone N, Patnaik JL, Rich E, Ahmed M, Husain L, Hassan MY, Tanjil MSH, Weitzman D, Dai T, Wagner BD, Cherwek DH, Congdon N, Islam K. Autonomous artificial intelligence increases real-world specialist clinic productivity in a cluster-randomized trial. *NPJ Digit Med*. 2023; 6(1): 184. doi: 10.1038/s41746-023-00931-7.
20. Nittas V, Daniore P, Landers C, Gille F, Amann J, Hubbs S, Puhon MA, Vayena E, Blasimme A. Beyond high hopes: A scoping review of the 2019-2021 scientific discourse on machine learning in medical imaging. *PLOS Digit Health*. 2023; 2(1): e0000189. doi: 10.1371/journal.pdig.0000189.
21. Zhou K, Liu Z, Qiao Y, Xiang T, Loy CC. Domain Generalization: A Survey. *IEEE Trans Pattern Anal Mach Intell*. 2023; 45(4): 4396-4415. doi: 10.1109/TPAMI.2022.3195549.
22. Benzinger L, Ursin F, Balke WT, Kacprowski T, Salloch S. Should Artificial Intelligence be used to support clinical ethical decision-making? A systematic review of reasons. *BMC Med Ethics*. 2023; 24(1): 48. doi: 10.1186/s12910-023-00929-6.
23. Jung J, Lee H, Jung H, Kim H. Essential properties and explanation effectiveness of explainable artificial intelligence in healthcare: A systematic review. *Heliyon*. 2023; 9(5): e16110. doi: 10.1016/j.heliyon.2023.e16110.
24. Kerasidou A. Ethics of artificial intelligence in global health: Explainability, algorithmic bias and trust. *J Oral Biol Craniofac Res*. 2021; 11(4): 612-614. doi: 10.1016/j.jobocr.2021.09.004.
25. Ouanes K, Farhah N. Effectiveness of Artificial Intelligence (AI) in Clinical Decision Support Systems and Care Delivery. *J Med Syst*. 2024; 48(1): 74. doi: 10.1007/s10916-024-02098-4.
26. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med*. 2019; 25(1): 44-56. doi: 10.1038/s41591-018-0300-7.
27. Ahmed MI, Spooner B, Isherwood J, Lane M, Orrock E, Dennison A. A Systematic Review of the Barriers to the Implementation of Artificial Intelligence in Healthcare. *Cureus*. 2023; 15(10): e46454. doi: 10.7759/cureus.46454.

ENDEREÇO CORRESPONDÊNCIA

GABRIEL RIBEIRO DE OLIVEIRA SILVA
BR 364, KM 195, No. 3800, Jataí, GO, Brasil
E-mail: oliveira.gabriel@discente.ufj.edu.br

EDITORIA E REVISÃO

Editores chefes:

Waldemar Naves do Amaral - <http://lattes.cnpq.br/4092560599116579> - <https://orcid.org/0000-0002-0824-1138>
Nílzio Antônio da Silva - <http://lattes.cnpq.br/1780564621664455> - <https://orcid.org/0000-0002-6133-0498>

Autores:

Gabriel Ribeiro de Oliveira Silva - <http://lattes.cnpq.br/9082313490110126> - <https://orcid.org/0009-0004-2846-762X>

Ítalo Bezerra Alves - <http://lattes.cnpq.br/5900325370274094> - <https://orcid.org/0009-0000-2393-3135>

Isabela Luiza Pacheco Rodrigues - <http://lattes.cnpq.br/7700244285082906> - <https://orcid.org/0009-0009-3936-0007>

Douglas Wallacy da Silva Amâncio - <http://lattes.cnpq.br/3261605374734058> - <https://orcid.org/0009-0003-2058-308X>

Fabiano Inácio de Souza - <http://lattes.cnpq.br/8305270225089632> - <https://orcid.org/0000-0001-5862-4674>

Revisão Bibliotecária: Izabella Goulart

Revisão Ortográfica: Dario Alvares

Tradução: Soledad Montalbetti Magri

Recebido: 24/08/26. Aceito: 27/08/26. Publicado em: 29/09/2026.